PATENT APPLICATION

SC
#36
5/24/04

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Applicants: Kephart et al.

Examiner: Bashore, William

Serial No: 09/061,706

Group: Art Unit 2176

Filed: April 17, 1998

Docket: YO998-143 (8728-503)

For: **An Automated Assistant for Organizing Electronic Documents**

APPEAL BRIEF

Appeal from Group 2176

F. Chau & Associates, LLC
1900 Hempstead Turnpike
East Meadow, New York 1553
TEL: (516) 357-0091
FAX: (516) 357-0092
Attorneys for Appellants

TABLE OF CONTENTS

I. INTRODUCTION

This Appeal is from a Final Office Action mailed on October 2, 2003 (Paper No. 31) (hereinafter, referred to as the "Final Action") finally rejecting claims 11, 13-21, 23-33 and 37-68 of the above-identified application, and an Advisory Action mailed on December 30, 2003 (Paper No. 33). Applicants commenced this Appeal by a Notice of Appeal filed on February 4, 2004, and hereby submit this Appeal Brief.

II. REAL PARTY IN INTEREST

The real party in interest for the above-identified application is International Business Machines (IBM) Corporation, the assignee of the entire right, title and interest in and to the subject application by virtue of an assignment of recorded in the U.S. Patent and Trademark Office at reel 9645 frame 0117.

III. RELATED APPEALS AND INTERFERENCES

There are no Appeals or Interferences known to Applicant, Applicant's representatives or the Assignee, which would directly affect or be indirectly affected by or have a bearing on the Board's decision in the pending Appeal.

IV. STATUS OF CLAIMS

Claims 11, 13-21, 23-33 and 37-68 are pending, stand rejected and are under appeal. The claims on appeal are set forth in the attached Appendix.

Claims 11 and 61 are independent claims. Claims 13-21, 23-33, 37-60, 62, 64-68 depend directly or indirectly from claim 11. Claim 63 depends from claim 61.

V. STATUS OF AMENDMENTS

No after final Amendments were filed in this case subsequent to the Final Action.

VI. SUMMARY OF THE INVENTION

In general, the claimed inventions are related to systems and methods for managing electronic documents such as electronic files, electronic mail (e-mail), web pages, etc. Systems and methods for document management according to the invention implement methods for automatically classifying electronic documents to provide assistance to a user for categorizing electronic documents for filing within a collection (e.g., directory, folder, etc). The classification process is trained to automatically determine the categories that are most likely to be selected by the user for categorizing a document. The classification process obtains the most likely categorization labels, which are presented to the user (via a graphical user interface) in a manner that accentuates such most likely categorization labels for fast and efficient labeling/categorizing of the electronic document by the user. The claimed inventions further include methods for incrementally retraining the classifier using a "lazy strategy", to adaptively learn to classify documents as modifications of the contents of the collections (e.g., folders) change over time.

More specifically, by way of example, the claimed inventions of claims 11 and 62 recite a method and a program storage device having executable instructions, respectively, for providing

an automated method of assisting a user with the task of categorizing electronic documents into a collection, comprising the steps of:

classifying, with a classifier, a document to obtain a plurality of most likely categorical labels;

deriving a plurality of categorizations shortcuts from the plurality of most likely categorical labels;

displaying, to the user, a representation of the plurality of most likely categorical labels;

receiving , from the user, a selection of one or more of the most likely categorical labels representative of the document to be categorized within a collection;

labeling the document within the collection with one or more of the selected categorical labels; and

incrementally retraining the classifier to adapt to modifications of the collection, wherein the incremental retraining is performed using a lazy strategy for incrementally retraining the classifier.

In one embodiment of the invention, the inventive method is imbedded in an electronic mail application to assist a user in organizing his/her mail into separate folders. In another embodiment, the inventive method is implemented in a Web browser to assist a user in organizing "bookmarks", which are pointers to web pages. For purposes of illustration, the inventions of claims 11 and 61 will be discussed hereafter with reference to the exemplary embodiments depicted in the Figures and the corresponding description in Applicants' specification (hereinafter, "Spec.") as related to managing electronic messages in an e-mail application. It is to be understood that the following description of the claimed inventions with

reference to e-mail documents is for illustrative purposes to provide some context for the claimed inventions, but nothing herein shall be construed as placing any limitation on the claimed inventions.

In the context of managing electronic messages, the inventions of claims 11 and 61 provide *an automated method to assist a user in categorizing electronic messages into a collection*, wherein folders provide a mechanism for categorizing the electronic messages so that all related messages can be stored in the same folder in a database (see, e.g., Spec., page 12, lines 6-11).

Further, with respect to the exemplary e-mail embodiment, the step of *classifying, with a classifier, a document to obtain a plurality of most likely categorical labels*, may comprise automatically classifying an electronic message using a text classifier (a classifier procedure, *Classsifeir_Classify*) to predict the most probable destination folders that a user may consider for storing the electronic message. In such instance, the *categorical labels* relate to the names of the most probable destination folders (see, e.g., Spec., page 12, lines 6-22; page 15, lines 2-6; and FIG. 3, step 308). The core classifier procedure Classifier_Classify takes an electronic message as its input and produces, as output, a list of one or more categorical labels or folder names. In the exemplary embodiment, the folders provide a mechanism for categorizing messages. For instance, the user may use the folder "Baseball" to store all messages related to the game of baseball.

Moreover, with respect to the exemplary e-mail embodiment, the recited steps of: *deriving a plurality of categorizations shortcuts from the plurality of most likely categorical labels; displaying, to the user, a representation of the plurality of most likely categorical labels;*

receiving, from the user, a selection of one or more of the most likely categorical labels representative of the document to be categorized within a collection; and labeling the document within the collection with one or more of the selected categorical labels, may include the example methods as depicted and described in the exemplary embodiments of FIGs., 2 and 3 and corresponding text (e.g., Spec., pages 12-16).

More specifically, by way of example, the *deriving* step may comprise labeling display buttons with the plurality of most likely categorical labels (folder names), wherein the displaying step comprises displaying the labeled display buttons with the electronic document (see, e.g., steps 310, 312, FIG. 3 of Spec.). More specifically, by way of example with reference to FIG. 2 of Spec., three (3) "Move To" buttons are displayed with the message, wherein the "Move To" buttons are each labeled with the name of one of the folders that were determined by the classifier to be the three (3) most likely selected by the user as the destination folder for the message (see, e.g., corresponding description in Spec., page 12, lines 18-22).

In another exemplary embodiment, the *deriving* step comprises creating an ordered set of the plurality of most likely categorical labels (folder names), and the displaying step comprises displaying with the document the ordered set prepended to a standard ordered set of other categorical labels (see, e.g., FIG. 3, steps 310, 318; page 15, line 20- page 16, line 1).

Thereafter, the user can select *one or more of the most likely categorical labels* (e.g., names of most likely destination folders) *and the electronic document will be labeled within the collection with one or more of the selected categorical labels* (e.g., storing the document in folders or locations of the collection corresponding to the one or more selected categorical labels) (see, e.g., Spec., FIG. 4, steps 314, 320, 324; page 16). For instance, with a "categorization

shortcut" represented as a displayed "Move To" button, the user can select (via mouse click) one of these buttons, in which instance the message is moved to the folder name displayed with the "Move To" shortcut button.

Further, with respect to the exemplary e-mail embodiment, the step of *incrementally retraining the classifier to adapt to modifications of the collection, wherein the incremental retraining is performed using a lazy strategy for incrementally retraining the classifier,* may comprise methods that enable updates (incremental learning) to the classifier to be "deferred" (lazy strategy) when the content (electronic messages) of a folder is modified (message added, deleted, moved). In particular, in the exemplary e-mail application, a user continually receives new mail, deletes old mail and moves mail messages among folders and, consequently, the contents of the folders are in constant flux. Therefore, an incremental learning process enables the classifier to continually adapt itself. Advantageously, using a lazy learning strategy makes it possible to perform the classifier update, for example, during a moment when that update is less likely to hurt performance, and the lazy learning strategy decreases the vigilance required of the classifier thereby decreasing the need for tight communication between the categorizer and the mail application, which can complicate the incorporation of automated categorization into an existing mail application (see, e.g., Spec., page 19, lines 6-20). A "lazy strategy" for incremental learning is contrasted with an "instant" strategy, in which updates are made to the classifier immediately whenever electronic messages are added, deleted or moved from the folders (see, e.g., Spec., page 18, line 16- page 20, line 5).

VII. ISSUES

(1) Claims 11, 13-21, 23-33, 44-45, 47-51, and 60-68 stand rejected under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent No. 5,544,360 to Lewak in view of U.S. Patent No. 6,029,195 to Herz.

Thus, one issue on appeal is whether the combination of Lewak and Herz is legally sufficient to establish a *prima facie* case of obviousness against claims 11, 13-21, 23-33, 44-45, 47-51, and 60-68.

(2) Claims 37-43 and 46 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Lewak in view of Herz in further view of U.S. Patent No. 5,867,799 to Lang et al.

Thus, another issue on appeal is whether the combination of Lewak, Herz and Lang is legally sufficient to establish a *prima facie* case of obviousness against claims 37-43 and 46.

(3) Claims 52 and 56-59 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Lewak in view of Herz in further view of Using Netscape, 1995 Que Corporation, pgs. 55 and 67.

Thus, another issue on appeal is whether the combination of Lewak, Herz and Netscape is legally sufficient to establish a *prima facie* case of obviousness against claims 52 and 56-59.

(4) Claims 53-55 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Lewak in view of Herz and Lang as applied to claims 26, 29, and 32 above, and in further view of Netscape.

Thus, another issue on appeal is whether the combination of Lewak, Herz, Lang, and Netscape is legally sufficient to establish a *prima facie* case of obviousness against claims 53-55.

VIII. GROUPING OF CLAIMS

For Issue (1) above:

(i) Claims 13-16, 17-21, 23-33, 44-45, 47-51, 60, 62, 64-66 and 67-68 stand or fall with Claim 11; and

(ii) Claim 63 stands or falls with Claim 61.

For Issue (2) above:

Claims 37-43 and 46 stand or fall with Claim 11.

For Issue (3) above:

Claims 52 and 56-59 stand or fall with Claim 11.

For Issue (4) above:

Claims 53-55 stand or fall with Claim 11.

IX. ARGUMENTS

A. The Combination of Lewak and Herz is Legally Deficient to Support a *Prima Facie* Case Of Obviousness Against Claims 11 and 61

In rejecting claims under 35 U.S.C. 103, the Examiner bears the initial burden of presenting a prima facie case of obviousness. In re Rijckaert, 9 F.3d 1531, 1532 (Fed. Cir. 1993). The burden of presenting a prima facie case of obviousness is only satisfied by showing some objective teaching in the prior art or that knowledge generally available to one of ordinary skill in the art would lead that individual to combine the relevant teachings of the references. In re Fine, 837 F.2d 1071, 1074 (Fed. Cir. 1988). A prima facie case of obviousness is established when the teachings of the prior art itself would appear to have suggested the claimed subject matter to

one of ordinary skill in the art. In re Bell, 991 F.2d 781, 782 (Fed. Cir. 1993). The suggestion to combine the references should come from the prior art, and the Examiner cannot use hindsight gleaned from the invention itself to pick and choose among related disclosures in the prior art to arrive at the claimed invention. In re Fine, 837 F.2d at 1075. If the Examiner fails to establish a prima facie case, the rejection is improper and must be overturned. In re Rijckaert, 9 F.3d at 1532 (citing In re Fine, 837 F.2d at 1074).

In the case at bar, claims 11 and 61 stand finally rejected as being obvious based on the combination of Lewak and Herz. The rejections of claims 11 and 61 are based primarily on the teachings of Lewak, with reliance on the teachings of Herz to cure the deficiencies of Lewak. For purposes a discussion, a brief summary of the relevant teachings of Lewak and Herz.

Lewak discloses a file management system, which implements a manual method for categorizing electronic documents via a GUI (graphic user interface). The GUI allows a user to define and edit categories for files, and manually associate files with particular categories for storage in a computer system (see, e.g., Col. 3, lines 60-67; Col. 4, line 66 – Col. 5, line 2).

Herz is directed to *information retrieval systems* and methods, which enable users to access target objects of relevance and interest to the users (e.g., searching and accessing information from on-line information sources over a data communications network) (Col. 4, lines 28-46). The target objects may be news, electronic mail, published documents, or product descriptions (Col. 6, lines 59-63). Herz teaches constructing a "target profile" for each target object based on descriptive attributes of the object, and constructing a "target profile interest summary" for a user, which consists of a summary of "target profiles" that the user likes/dislikes or a plurality of search profiles for different topics. Further, Herz teaches a method for

searching/accessing target objects by comparing the "target profiles" of "target objects" with the "search profiles" of the user to estimate the user's interest in various target objects and generate for the user a customized rank-ordered listing of target objects most likely to be of interest to the user. The user's target profile interest summary is automatically updated on a continuing basis to reflect the user's changing interests (in Col. 4, lines 48-62; Col. 7, lines 1-18).

Appellants respectfully submit that at the very least, the combination of Lewak and Herz is legally deficient to support a *prima facie* case of obviousness against claims 11 and 61. In particular, Appellants will demonstrate that the combination of Lewak and Herz fails to disclose or suggest various elements of claims 11 and 61. Moreover, Appellants will demonstrate that no reasonable, objective basis has been articulated in the Final Action to support the proposed combination of Lewak and Herz and that one of ordinary skill in the art would not be motivated to combine the relevant teachings of Lewak and Herz in the manner suggested in the Final Action. In fact, it will become readily apparent that the obviousness rejections are legally and logically flawed, being primarily grounded on hindsight reasoning in view of Applicants' specification, as well as strained interpretations and mischaracterizations of the teachings of Lewak and Herz , in an attempt to derive the claimed inventions.

The Examiner's basis for rejecting claim 11 is set forth on pages 3-5 of the Final Action. Further, the basis for rejecting claim 61 is set forth on pages 11-13 of the Final Action. In light of the similarity of the claimed subject matter of claims 11 and 61, the obviousness arguments for claims 11 and 61 are essentially identical. As such, Appellants will address the impropriety of the obviousness rejections with specific reference to the arguments set forth for claim 11 on pages 3-5 of the Final Action, which is equally applicable to the rejection of claim 61.

(i) The combination of Lewak and Herz does not disclose or suggest *classifying, with a classifier, a document to obtain a plurality of most likely categorical labels.*

Examiner's obvious analysis of claim 11 begins on page 3 of the Final Action by citing Lewak (Col. 8, lines 6-15; Col. 9, lines 50-55) as disclosing ***classifying, with a classifier, a document ....*** However, the Examiner acknowledges that Lewak does not specifically teach ***classifying, with a classifier, a document to obtain a plurality of most likely categorical labels*** .

In an effort to cure the deficiencies of Lewak in this regard, the Examiner relies on Herz (Col. 7, lines 10-16) as teaching "**a method of generating for each user a customized rank-ordered listing of target objects most likely to be of interest to that user**". The Examiner attempts to analogize this teaching of Herz to the claim element ***... to obtain a plurality of most likely categorical labels***, contending that it would have been obvious to apply the method of Herz to the method of Lewak because of Herz's "taught advantage of document suggestion, providing an efficient method of allowing users to select articles of inter from a large set of articles (Col. 2, lines 40-42)".

It is respectfully submitted that Examiner's reliance on Herz in this regard is wholly misplaced. Indeed, Examiner focuses too narrowly on specific "keywords" and "catch phrases" without due consideration to the relevant contexts of the claimed invention and the teachings of Herz, which results in a strained, erroneous interpretation.

In particular, as explained above, Herz is directed to information retrieval systems and methods, which enables a user to access "target objects" (e.g., electronic documents) of relevance and interest to the user. Herz teaches a method for searching/accessing target objects by

comparing the "target profiles" of "target objects" with the "search profiles" of the user to estimate the user's interest in various "target objects" and generate for the user a customized rank-ordered listing of "target objects" most likely to be of interest to the user.

In stark contrast, with the claimed invention, a document is classified (via a classifier method) to obtain a listing of the most likely categorical labels that the user would use to categorize the document.

These methods are fundamentally different – the Herz method, as a result of a search process, generates a list of "desirable" target objects (e.g., desirable electronic documents) for access by the user, whereas in the claimed invention, an electronic document is classified to obtain a plurality of "desirable" categorical labels for the categorizing the electronic document.

In this regard, the relevant teachings of Lewak and Herz, as cited in the Final Action, are so fundamentally distinct in terms of both functionality and purpose, that one or ordinary skill in the art would not be motivated to combine the teachings in the manner suggested. Most tellingly, Examiner's conclusion of obviousness, which is based on "Herz's taught advantage of document suggestion, providing an efficient method of allowing users to select articles of interest from a large set of articles" as motivation for combining Herz and Lewak, is clearly erroneous. Indeed, "Herz's taught advantage of **document suggestion**" appears to be wholly irrelevant to the claimed method of classifying a document for **suggesting categories** for categorizing an electronic document. Accordingly, it is apparent that Examiner's grounds of obviousness in view of Herz's teaching of a "ranked order listing of target objects" is nothing more than a textbook example of impermissible hindsight reasoning based on the teachings of Appellants specification.

(ii) The combination of Lewak and Herz does not disclose or suggest *deriving a plurality of categorizations shortcuts from the plurality of most likely categorical labels*

On page 3 of the Final Action, Examiner asserts that Lewak's teaching of "linking category descriptions" (Col. 15, lines 39-46) reasonably suggests "*deriving a plurality of categorizations shortcuts* ..." as claimed in claim 11. It is respectfully submitted however, that such assertion is premised on a strained interpretation of the teachings of Lewak as applied to the claim element.

Lewak discloses (in Col. 15, lines 28-55) a "linking category" description, which is a special category that provides a mechanism to "link" related/corresponding category descriptions. A linking category description (e.g., E-MAIL) can be "linked" to other linked category descriptions (e.g., Sent, Received, Action, ...) so when a user categorizes a file with one category description, the user would be given an indication that the file should also be categorized by a related category description. These "links" are handles that are specified in the data structures of the category descriptions (see, Col. 5, line 39, - Col. 6, line12).

Lewak's teaching of "linking category descriptions" does not disclose or suggest "characterization shortcuts" as contemplated by the claimed inventions. Indeed, in the first instance, it is to be noted that the Examiner's analysis conveniently ignores the full scope of the claim language, which reads *deriving a plurality of categorizations shortcuts from the plurality of most likely categorical labels*. As noted above in the Summary (section VI), in the context of the claimed invention, "categorization shortcuts" provide methods for presenting the most likely categorical labels (as determined by classification) in manner that accentuates such labels for fast and efficient labeling/categorizing of the electronic document.

In other words, in the context of the claimed inventions, "categorization shortcuts" are not mechanisms for "linking" related category descriptions, but rather mechanisms that enable easy identification and selection of a "most likely category" for an electronic document. In fact, the Examiner appears to have narrowly focused on the term "link" and "shortcut" without due consideration of the proper claim context.

The erroneousness of such analysis is further evident from Examiner's basis of "reasonable suggestion" (as set forth on the bottom of Page 3 of the Final Action) that the "linked descriptions ... , which help to provide freedom from the restrictions imposed by present day computer filing systems", citing Lewak, Col. 3, lines 62-65." This basis of "reasonable suggestion" appears to be irrelevant at best, as it clearly does not explain how "linking category descriptions" even remotely suggests the claimed element of *deriving a plurality of categorizations shortcuts from the plurality of most likely categorical labels*.

(iii) The combination of Lewak and Herz does not disclose or suggest *incrementally retraining the classifier to adapt to modifications of the collection, wherein the incremental retraining is performed using a lazy strategy*

On page 4 of the Final Action, the Examiner attempts to demonstrate how the combined teachings of Lewak and Herz either disclose or suggest the claimed features of *incrementally retraining the classifier to adapt to modifications of the collection, wherein the incremental retraining is performed using a lazy strategy,* as recited in claim 11. However, as will become readily apparent, the Examiner's analysis in this regard is fundamentally flawed, both logically and legally.

Indeed, in the first instance, the analysis is based, in part, on contradictory assertions. In particular, on one hand, the Examiner acknowledges that "Lewak does not specifically teach a method of incrementally retraining a classifier". But on the other hand, the Examiner relies on Lewak (Col. 7, lines 50-67) as essentially providing support for establishing "*wherein the incremental retraining is performed using a lazy strategy for incrementally retraining the classifier*" based on Lewak's purported teaching of "deferred categorization" (which Examiner characterizes as process whereby a FC (file control) manager determines (during periods of inactivity) those files that have not been categorized, and notifies the user of such "uncategorized" files).

To the extent that Examiner acknowledges that Lewak does not teach *incrementally retraining a classifier*, it necessarily and logically follows that Lewak cannot teach or suggest *incrementally retraining a classifier using a lazy strategy.* Accordingly, these contradictory contentions provide a *prima facie* indication of the erroneousness of the analysis. At the very least, such conclusion can be viewed as nothing more than selective combination of out-of-context keywords and improper hindsight reasoning to derive the claim.

In any event, the Examiner's analysis with regard to "incremental training" is based on gross mischaracterizations of the teachings of Lewak and Herz as applied to the claim elements. For instance, Examiner's analogy between Herz's teaching of ***automatically updating a user target profile interest summary on a continuing basis*** (Col. 7, lines 15-17) and the claimed element of ***incrementally retraining a classifier to adapt to modifications of the collection*** , is wholly erroneous. Indeed, there is simply no correlation between the "user target profile interest summary" and a "classifier".

In particular, the *user target profile interest summary* is nothing more than a set of attributes that specifies user interests, which is input to *a profile processing module* (along with *target object profiles)* to identify target objects most likely to be of interest to the user (see, Col. 7, lines 4-16; Col. 4, lines 48-64).

In stark contrast, in the context of the claimed inventions, a "classifier" is an automated process that is trained to automatically identify most likely categorization labels for categorizing an electronic document. The *user target profile interest summary (*which comprises a list of user attributes) is not even remotely the same as a classifier (automated process). In this regard, *updating the user target profile interest summary* has no relation, whatsoever, to *incrementally retraining the classifier*, as claimed in claims 11 and 61.

Moreover, Examiner's reliance on Lewak's teaching (Col. 7, lines 50-67) of "deferred categorization" as disclosing or suggesting "incremental retraining a classifier using a lazy strategy", is glaringly erroneous in that there is simply no relation between "deferred categorization" and "incrementally retraining a classifier, much less incrementally retraining using a lazy strategy", as essentially claimed. In particular, as noted above (and as articulated on pages 4-5 of the Final Action) the Examiner characterizes "deferred categorization" as process whereby a FC (file control) manager determines (during periods of inactivity) those files that have not been categorized, and notifies the user of such "uncategorized" files. Even assuming, *arguendo*, that Examiner's characterization of Lewak as somehow teaching a "lazy strategy" for categorizing a document is correct, such teaching clearly provides no basis, whatsoever, for teaching or suggesting incrementally training a classifier using a lazy (deferred strategy). Indeed, on a fundamental level, a deferred process of retraining an automated method that learns to

automatically classify a document is entirely different from a deferred process for categorizing a document.

B. CONCLUSION

Accordingly, for at least the reasons set forth above, claims 11 and 61 are patentable and nonobvious over the combination of Lewak and Herz. Moreover, given that all remaining obviousness rejections are based entirely, or in part, on the combination of Lewak and Herz, as applied to base claims 11 and 61, the obviousness rejections for all dependent claims are legally deficient at least for the same reasons given for claim 11 and 61. Therefore, it is respectfully requested that the Board reverse all claim rejections under 35 U.S.C. § 103(a).

Respectfully submitted,

Frank V. DeRosa
Reg. No. 43,584
Attorney for Applicant(s)

F. Chau & Associates, LLC
1900 Hempstead Turnpike
East Meadow, New York 11553
TEL: (516) 357-0091
FAX: (516) 357-0092

APPENDIX A

1. (Canceled).

2. (Canceled).

3. (Canceled).

4. (Canceled).

5. (Canceled).

6. (Canceled).

7. (Canceled).

8. (Canceled).

9. (Canceled).

10. (Canceled).

11. An automated method of assisting a user with the task of categorizing electronic documents into a collection, comprising the steps of:

classifying, with a classifier, a document to obtain a plurality of most likely categorical labels;

deriving a plurality of categorizations shortcuts from the plurality of most likely categorical labels;

displaying, to the user, a representation of the plurality of most likely categorical labels;

receiving , from the user, a selection of one or more of the most likely categorical labels representative of the document to be categorized within a collection;

labeling the document within the collection with one or more of the selected categorical

labels; and

incrementally retraining the classifier to adapt to modifications of the collection, wherein the incremental retraining is performed using a lazy strategy for incrementally retraining the classifier.

12 (Canceled).

13. The method of claim 11 wherein the classifying step comprises the step of classifying, upon receipt into data storage, the document to obtain the plurality of most likely categorical labels.

14. The method of claim 11 wherein the deriving step comprises the step of deriving, upon receipt of the document into data storage, categorization shortcuts from the plurality of most likely categorical labels.

15. The method of claim 11 wherein the deriving step comprises the step of labeling display buttons with the plurality of most likely categorical labels, and the displaying step comprises the step of displaying the labeled display buttons with the document.

16. The method of claim 11 wherein the deriving step comprises the step of creating an ordered set of the plurality of most likely categorical labels, and the displaying step comprises the step of displaying with the document the ordered set prepended to a standard ordered set of

other categorical labels.

17. The method of claim 11 wherein the classifying step occurs substantially simultaneously with the displaying step.

18. The method of claim 11 wherein the classifying step comprises the step of classifying upon invocation by the user, the document to obtain the plurality of most likely categorical labels.

19. The method of claim 18 wherein the invocation comprises a selection by the user of a classify button.

20. The method of claim 11 wherein the labeling step comprises the step of storing the document in folders or locations of the collection corresponding to the one or more selected categorical labels.

21. The method of claim 11 further comprising the step of displaying a standard list of all categorical labels, wherein the receiving step comprises the step of receiving, from the user, data representative of one or more selected categorical labels from either the plurality of displayed categorization shortcuts or the standard list.

22 (Canceled).

23. The method of claim 11 wherein the retraining step comprises the step of retraining the classifier in response to the labeling step.

24. The method of claim 11 wherein the labeling step comprises the step of storing the document in folders or locations of the collection corresponding to the one or more selected categorical labels and the retraining step comprises the steps of:

receiving, from the user, addition data representative of an addition of a document into a tofolder; and

retraining the classifier in response to the addition data.

25. The method of claim 24 wherein the retraining step comprises the step of assigning, in the classifier, the added document to the tofolder.

26. The method of claim 25 further comprising the step of identifying excluded folders to be excluded from retraining and wherein the retraining step comprises the step of assigning, in the classifier, the added document when the tofolder is not one of the identified excluded folders.

27. The method of claim 11 wherein the labeling step comprises the step of storing the document in folders or locations of the collection corresponding to the one or more selected categorical labels and the retraining step comprises the steps of:

receiving, from the user, deletion data representative of a removal of a document from a from folder; and

retraining the classifier in response to the deletion data.

28. The method of claim 27 wherein the retraining step comprises the step of unassigning, in the classifier, the removed document from the from folder in which it was categorized.

29. The method of claim 28 further comprising the step of identifying excluded folders to be excluded from retraining and wherein the retraining step comprises the step of unassigning, in the classifier, the removed document when the from folder is not one of the identified excluded folders.

30. The method of claim 11 wherein the labeling step comprises the step of storing the document in folders or locations of the collection corresponding to the one or more selected categorical labels and the retraining step comprises the steps of:

receiving, from the user, move data representative of a movement of a document from a source folder to a destination folder; and

retraining the classifier in response to the move data.

31. The method of claim 30 wherein the retraining step comprises the steps of:

unassigning, the classifier, the moved document from the source folder in

which it was categorized; and

assigning, in the classifier, the moved document to the destination folder.

32. The method of claim 31 further comprising the step of identifying excluded folders to be excluded from retraining and wherein the retraining step comprises the steps of:

unassigning, in the classifier, the moved document when the source folder is not one of the identified excluded folders; and

assigning, in the classifier, the moved document when the destination folder is not one of the identified excluded folders.

33. The method of claim 11 wherein the retraining step occurs immediately after a collection modification.

34. (Canceled).

35. (Canceled).

36. (Canceled).

37. The method of claim 20, wherein the classifying step comprises the steps of :

tokenizing the document into different tokens;

tallying a number of occurrences of each token in the document;

computing, for each folder, a token weight of each token;

comparing, for each token, the number of occurrences and the token weights;

creating a similarity score in response to the comparing step; and

identifying a subset of folders for which the similarity score is highest.

38. The method of claim 37 further comprising the step of removing, from the identified subset, all folders from which the similarity score is lower than a default or specified threshold.

39. The method of claim 37, wherein the computing step comprises the step of computing the token counts of each token in each of the folders.

40. The method of claim 37 wherein the tokenizing step comprises the steps of:

separately tokenizing different portions of the document; and

labeling the tokens according to the different portions;

41. The method of claim 25 wherein the classifying step comprises the steps of:

tokenizing the document into different tokens;

tallying a number of occurrences of each token in the document;

retrieving, for each folder, a tokencount of each token;

computing, for each folder, a token weight of each token;

comparing, for each token, the number of occurrences and the token weights;

creating a similarity score in response to the comparing step; and

identifying a subset of folders for which the similarity score is highest, and wherein the assigning step comprises the step of adding the number of occurrences of each token to the tokencount of the tofolder.

42. The method of claim 28 wherein the classifying step comprises the steps of :

tokenizing the document into different tokens;

tallying a number of occurrences of each token in the document;

retrieving, for each folder, a tokencount of each token;

computing, for each folder, a token weight of each token;

comparing, for each token, the number of occurrences and the token weights;

creating a similarity score in response to the comparing step; and

identifying a subset of folders for which the similarity score is highest, and wherein the unassigning step comprises the step of subtracting the number of occurrences of each token to the tokencount of the tofolder.

43. The method of claim 31 wherein the classifying steps comprises the steps of:

tokenizing the document into different tokens;

tallying a number of occurrences of each token in the document;

retrieving, for each folder, a tokencount of each token;

computing, for each folder, a token weight of each token;

comparing, for each token, the number of occurrences and the token weights;

creating a similarity score in response to the comparing step; and

identifying a subset of folders for which the similarity score is highest, and

wherein the unassigning step comprises the step of subtracting the number of occurrences of each token from the tokencount of the source folder, and the assigning step comprises the step of adding the number of occurrences of each token to the tokencount of the destination folder.

44. The method of claim 11, further comprising the step of training the classifier from scratch with a pre-existing collection of categorized documents.

45. The method of claim 44 wherein the labeling step comprises the step of storing the document in folders or locations of the collection corresponding to the one or more selected categorical labels and the training step comprises the step of assigning, in the classifier, each of the pre-existing documents to a folder in which it is categorized.

46. The method of claim 45 wherein the classifying step comprises the steps of:

tokenizing the document into different tokens;

tallying a number of occurrences of each token in the document;

retrieving, for each folder, a tokencount of each token;

computing, for each folder, a token weight of each token;

comparing, for each token, the number of occurrences and the token weight;

creating a similarity score in response to the comparing step; and

identifying a subset of folders for which the similarity score is highest, and

wherein the assigning step comprises the step of adding the number of occurrences of each token to the tokencount of the tofolder.

47. The method of claim 45 further comprising the step of identifying excluded folders to be excluded from training and wherein the training step comprises the step of assigning, in the classifier, each of the pre-existing documents, except those in the identified excluded folders.

48. The method of claim 11 wherein the labeling step comprises the step of storing the document in folders or locations of the collection corresponding to the one or more selected categorical labels and the re-training step comprises the steps of:

determining a time of a last step of retraining; and

retraining the classifier on each folder which was modified after the determined time.

49. The method of claim 11 wherein labeling step comprises the step of storing the document in folders or locations of the collection corresponding to the one or more selected categorical labels, the method further comprising the step of training the classifier from scratch with a pre-existing collection of categorized documents, wherein the retraining step comprises the steps of:

determining a time of the step of training or a last step of re-training; and

retraining the classifier on each folder which was modified after the determined

time.

50. The method of claim 11 wherein the classifying step uses the TF-IDF principle.

51. The method of claim 11 wherein the electronic document is an e-mail message.

52. The method of claim 11 wherein the electronic document is web page and the collection is a collection of bookmarks.

53. The method of claim 41 wherein the electronic document is web page and the collection is a collection of bookmarks, the method further comprising the step of storing, for each web page, a pagetokencount matching the tallied number of occurrences of each token.

54. The method of claim 42 wherein the electronic document is a web page and the collection is a collection of bookmarks, the method further comprising the step of storing, for each web page, a pagetokencount matching the tallied number of occurrences of each token, wherein the unassigning step comprises the step of subtracting the pagetokencount from the tokencount of the fromfolder.

55. The method of claim 43 wherein the electronic document is a web page and the collection is a collection of bookmarks, the method further comprising the step of storing, for each web page, a pagetokencount matching the tallied number of occurrences of each token,

wherein the unassigning step comprises the step of subtracting the pagetokencount from the tokencount of the fromfolder.

56. The method of claim 11 wherein the electronic document is a multimedia document.

57. The method of claim 56 wherein the multimedia document is an image file, a video file or an audio file.

58. The method of claim 56 wherein the multimedia document combines any combination of text, an image file, a video file and an audio file.

59. The method of claim 57 wherein the multimedia document further includes text.

60. The method of claim 11 wherein the electronic document comprises data sets that are not viewable in their entirety, but can be categorized in response to some presentation to the user.

61. A program storage device, readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for automatically assisting a user with the task of categorizing an electronic document into a collection, the method comprising the steps of:

classifying, with a classifier, a document to obtain a plurality of most likely categorical labels;

deriving a plurality of categorizations shortcuts from the plurality of most likely categorical labels;

displaying, to the user, a representation of the plurality of most likely categorical labels;

receiving , from the user, a selection of one or more of the most likely categorical labels representative of the document to be categorized within a collection;

labeling the document within the collection with one or more of the selected categorical labels; and

incrementally retraining the classifier to adapt to modifications of the collection, wherein the incremental retraining is performed using a lazy strategy for incrementally retraining the classifier.

62. The method of claim 11 wherein:

the step of displaying most likely categorical labels to the user further comprises displaying through a graphical user interface a plurality of virtual category buttons, each labeled with one of the categorical labels;

the step of receiving representative data from the user comprises receiving information that one of the category buttons has been selected by the user by clicking on the category button; and

the step of labeling the document comprises virtually moving the document to a virtual file folder unique to the categorical label selected, and occurs upon the clicking of the selected

category button without need for any other activity by the user.

63. The program storage device of claim 61 wherein:

the step of displaying most likely categorical labels to the user further comprises displaying through a graphical user interface a plurality of virtual category buttons, each labeled with one of the categorical labels;

the step of receiving representative data from the user comprises receiving information that one of the category buttons has been selected by the user by clicking on the category button; and

the step of labeling the document comprises virtually moving the document to virtual file folder unique to the categorical label selected, and occurs upon the clicking of the selected category button without need for any other activity by the user.

64. The automated method of claim 11, wherein the incrementally retraining is performed using a lazy strategy for incrementally retraining the classifier comprises:

deferring retraining of the classifier;

performing bookkeeping operations whenever messages are added to folders, removed from folders, or moved from one folder to another; and

automatically triggering retraining of the classifier by a predetermined criteria, wherein the retraining the classifier by processing any updates that have been deferred.

65. The automatic method of claim 64, wherein the predetermined criteria for

automatically performing the retraining step is a fixed amount of time that has elapsed since a last retraining step has been performed.

66. The automatic method of claim 64, wherein the predetermined criteria for automatically performing the retraining step is when a threshold number of documents have been added, deleted, or moved in the collection and any combination thereof.

67. The automatic method of claim 64, wherein the predetermined criteria for automatically performing the retraining step is when the system has reached an idle state or is either updating the classifier or performing bookkeeping operations whenever messages are added to folders, removed from folders, or moved from one folder to another.

68. The automatic method of claim 11, wherein incremental retraining the classifier includes an instant strategy for incrementally retraining the classifier.